

Revisiting the joint effect of temperature and relative humidity on airborne mold and bacteria concentration in indoor environment: A machine learning approach

Doheon Kim^a, Dongmin Shin^a, Dohyeong Kim^b , Boyeon Kwon^c , Choongki Min^c, Gloria Geevarghese^b, Seunghyun Kim^d, Jungho Hwang^{a,*}, SungChul Seo^{c,*}

^a Department of Mechanical Engineering, Yonsei University, Seoul, 03722, Republic of Korea

^b School of Economic, Political and Policy Sciences, University of Texas at Dallas, TX, 75080, USA

^c Department of Nano, Chemical and Biological Engineering, Seokyeong University, Seoul, 02713, Republic of Korea

^d Allergy Immunology Center, College of Medicine, Korea University, Seoul, 02841, Republic of Korea

ARTICLE INFO

Keywords:

Mold
Bacteria
Bioaerosol
Temperature
Relative humidity
Joint effects

ABSTRACTS

Exposure to airborne bioaerosols, such as bacteria and fungi, presents significant health risks, especially for vulnerable populations like children, the elderly, and those with compromised immune systems. Bioaerosol exposure can aggravate respiratory and allergic conditions, underscoring the need for real-time monitoring in indoor environments. However, continuous monitoring is often hindered by technical and economic challenges. Temperature and relative humidity are known to strongly influence airborne bacteria and mold levels, yet the combined impact of these factors remains insufficiently explored for effective control strategies in diverse indoor settings. This study investigates the joint effects of temperature and humidity on indoor bioaerosol concentrations through a machine learning-based outlier removal. Data were collected from 4,048 samples across 10 different types of multi-use facilities (e.g., daycare center, library) in South Korea, taking seasonal variations into account. A Random Forest model was employed to manage complex nonlinear relationships, identifying and excluding data points heavily influenced by external variables, such as ventilation and occupancy, rather than by temperature and humidity alone. The refined dataset enabled a focused analysis of temperature and humidity's combined impact, revealing specific conditions that correspond with elevated bioaerosol levels. The findings show that temperature and humidity jointly and significantly affect concentrations of bacteria and mold, with variations observed according to season and facility type. This research provides practical guidelines for controlling indoor bioaerosol levels by adjusting temperature and humidity alone, thus supporting safer and healthier indoor environments across a range of facilities.

1. Introduction

Humans suffer from allergies and respiratory diseases due to exposure to bioaerosols, particularly airborne bacteria and mold, which incurs significant social and economic costs [1,2]. Allergic rhinitis affects between 10 % and 30 % of the global population [3], and asthma affected over 250 million people and caused 455,000 deaths in 2019 [4]. Bioaerosols tend to be more prevalent and hazardous in indoor environments where people typically spend most of their time [5,6]. In particular, vulnerable groups, such as children, the elderly, and pregnant women, face greater risks from bioaerosol exposure due to their

weaker immune systems [7,8]. In South Korea, the prevalence of allergic rhinitis and atopic dermatitis among school-aged children and the elderly significantly increased from 2008 to 2017 [9], and the rates of asthma and severe asthma have also steadily risen over the past decade [10]. Thus, continuous monitoring and management of airborne microorganisms in indoor facilities, such as residential buildings, schools, offices, hospitals, and other enclosed spaces where people spend significant amounts of time, should play a key role to prevent the onset and worsening of asthmatic and allergic symptoms [8,11,12].

However, measuring the concentration of bioaerosols takes time due to the cultivation process, which makes continuous monitoring and

* Corresponding authors.

E-mail addresses: hwangjh@yonsei.ac.kr (J. Hwang), haha0694@skuniv.ac.kr (S. Seo).

<https://doi.org/10.1016/j.buildenv.2025.112548>

Received 11 November 2024; Received in revised form 5 January 2025; Accepted 7 January 2025

Available online 11 January 2025

0360-1323/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

rapid identification of bioaerosols in indoor air challenging. Various research efforts have been undertaken to develop real-time bioaerosol detection devices, including ATP bioluminescence methods [13,14], polymerase chain reaction (PCR) techniques [15], and antigen-antibody reaction approaches [16]. However, commercialization and implementation of a new monitoring device tends to face economic and technical issues, such as effective deployment of the device in indoor facilities and excessive cost of constant monitoring. This is a common concern across environmental and medical monitoring processes.

Recent advancements in computing technology position data-driven analytic methods as a promising solution to overcome the drawbacks of conventional bioaerosol monitoring and analysis tools [17–19]. Neural network-based deep learning techniques offer the potential to overcome the technical and practical limitations of traditional methods for measuring and estimating indoor bioaerosol concentrations in real-world environments. By identifying unrecognized outliers and accounting for the complex interactions among various environmental factors, AI-based models can improve indoor bioaerosol measurement and control by isolating the influence of individual factors, leading to more precise analysis and regulation [20–22].

Understanding the impact of environmental conditions on indoor bioaerosols is essential for effective air quality management [23]. Existing laboratory studies have revealed substantial linear or nonlinear correlations between bioaerosol concentrations and both temperature and relative humidity [24,25], as these factors are closely linked to bioaerosol viability [26–29]. Temperature affects the metabolic activity of bioaerosols and can alter their cell walls, with optimal survival typically observed in the 20–30 °C range. As for humidity, the hydrophilic groups on bioaerosol cell walls cause them to exist in droplet form upon emission. The evaporation rate of these droplets is influenced by humidity levels, potentially leading to dehydration or moisture-induced stress, which can reduce bioaerosol viability [26,30]. Importantly, the effects of temperature and humidity are interdependent rather than independent, highlighting the need for a comprehensive investigation into their combined impact on bioaerosol viability. Several laboratory studies have demonstrated a complementary relationship between temperature and relative humidity in influencing bioaerosol concentrations [31], emphasizing their interaction in affecting the survival and growth of microorganisms. Laboratory studies have shown that airborne bacteria and mold in indoor environments exhibit independent, combined, and interactive effects of temperature and humidity. However, field-based studies exploring the combined and interactive effects of temperature and humidity on airborne bacteria and mold in real-world settings are scarce. In contrast, there are numerous studies focusing on bacteria and mold on materials rather than in the air [32,33].

Previous field-based studies have evaluated independent linear correlations between temperature and relative humidity—each measured within limited temperature-humidity ranges and datasets—and indoor bioaerosol concentrations using Pearson correlation coefficients (R). These studies generally reported low correlations, with R values ranging from −0.2 to 0.4 [34–37]. Because Pearson correlation coefficients can only capture the linear effects of temperature and relative humidity individually on bioaerosol concentrations, they fail to account for the combined effects and nonlinear relationships. Moreover, unlike the controlled conditions of laboratory experiments, real indoor environments are subject to various uncontrollable factors, such as occupancy levels, ventilation, and facility usage [38,39]. It is crucial to investigate the optimal combination of temperature and humidity in diverse indoor settings that can minimize bioaerosol concentrations while accounting for various uncontrollable factors. This research should inform the development of effective guidelines for indoor environmental management focused on temperature and humidity control, but such efforts have been limited [12].

This study aims to analyze the combined effects of indoor temperature and relative humidity on airborne bacteria and mold concentrations in real-world indoor environments, identifying conditions that lead to

elevated levels. To achieve this, data from various indoor facilities were collected, and the relationship between airborne microbial concentrations and temperature and humidity was statistically examined. To separate the effects of only temperature and humidity on airborne bacteria and mold concentrations, a machine learning algorithm was developed to isolate the independent effects of temperature and humidity by filtering out data where other factors, such as occupancy, ventilation, and facility usage, were dominant influences. Using the refined dataset, the study visually mapped the temperature and humidity combinations associated with the unsafe level of bioaerosol concentrations for each facility type, providing specific and practical guidelines for effectively managing indoor bioaerosol levels.

2. Method

2.1. Data collection

We collected temperature, relative humidity, and airborne bacteria/mold concentrations from 4048 indoor locations in 10 different types of multi-use facilities in three regions of South Korea (Seoul, Gyeonggi and Chungcheong) from August 2021 to July 2023 across all seasons. Our method for measuring airborne bacteria and mold concentrations followed the Korean Indoor Air Quality Standard Method. Using a MAS-100eco sampler, airborne microorganisms (total airborne bacteria and mold) were collected at a height of 1.2–1.5 m from the floor, while temperature and humidity were measured simultaneously. Sampling was conducted under normal operating conditions with official approval obtained for sensitive facilities, such as daycare centers and healthcare institutions, as part of a project funded by the Korean Ministry of Environment. Contextual information, including the number of occupants, time of measurement (morning/afternoon), and the specific location within the facility, was recorded during sampling. Each data point represented the mean bioaerosol concentration calculated from three repeated measurements, with two data points collected per facility daily (morning and afternoon). Sampling extended from a minimum of one year to two years, encompassing seasonal variations—spring (March–May), summer (June–August), fall (September–November), and winter (December–February). To enhance representativeness, data were collected from a minimum of four and up to 28 facilities within each category, ensuring a robust and reliable dataset. To prevent contamination of the sampler and ensure measurement quality, the inside of the impactor was disinfected with 70 % alcohol swabs before and after sampling, and sterilized culture plates were used. TSA media were used for bacteria and PDA media for mold, with sampling conducted at a flow rate of 100 L/min for 1 min and 30 s. To determine the concentration, the collected bacteria were cultured at 35 °C ± 1 °C for 48 h, and the mold at 25 °C ± 1 °C for 5 to 7 days. The concentration of bacteria and mold (CFU/m³) was calculated by counting the colony-forming units (CFU) based on their growth.

The risk level for airborne microbial concentrations in indoor air, as defined by the Korean Ministry of Environment, is set at over 800 CFU/m³ for bacteria and 500 CFU/m³ for mold. These thresholds are based on guidelines from the WHO and standards from countries with indoor air quality management regulations. In this study, to propose conservative management standards for indoor airborne microbes, we designated 80 % of the risk level as the 'Caution' threshold. By using the 'Caution' threshold for binary classification, this study enables a proactive approach to indoor air quality management, allowing issues to be detected and addressed before reaching 'Danger' levels. Additionally, classifying based on the 'Caution' level increases the robustness of the model by incorporating more data, enhancing the reliability of early warning systems and improving overall management effectiveness.

2.2. Data preprocessing & augmentation

The collected data included various features, such as categorical

variables like facility type. To integrate these categorical variables into our artificial neural network (ANN) model, we employed 'Label Encoding,' where each category was assigned a unique integer. For example, facility types were converted into numerical values, with each distinct category represented by a specific integer. This allowed the ANN to process non-numeric inputs effectively. Additionally, instances exceeding the more conservative 'Caution' threshold were labeled as 1, while those within the 'Safety' criteria were labeled as 0, creating a clear binary distinction for analysis. To ensure consistency, normalization was applied to scale all data between 0 and 1. The normalized dataset was then split into training, validation, and test sets in a 7:1.5:1.5 ratio. This random partitioning ensured sample diversity, facilitating the development of a predictive model that remains robust under various conditions. The training set was used for model development, the validation set for hyperparameter tuning, and the test set for final performance evaluation.

The class imbalance between 'Safety' and 'Caution-Danger' posed a risk of model bias toward the majority class, potentially undermining overall performance. To mitigate this issue and improve model accuracy, we applied Gaussian noise-based oversampling, effectively balancing the data across both classes [40–42]. This data augmentation was applied exclusively to the training dataset. Based on the characteristics of each dataset, 3 % Gaussian noise was added to the sensor measurements for temperature and humidity, while 30 % and 25 % Gaussian noise was introduced to the concentrations of indoor airborne bacteria and mold, respectively.

2.3. Random forest model

We utilized a Random Forest (RF) model to classify the binary outcome of elevated bioaerosol levels, determining whether they exceeded the 'Caution' threshold. The classification was based mostly on label-encoded features, including facility type, indoor temperature, and humidity [43]. RF is a widely used machine learning method and a powerful algorithm that improves prediction performance by combining multiple decision trees [44]. The key components of RF are bagging and random subspace. Bagging generates multiple training datasets through bootstrap sampling, allowing each dataset to train a decision tree. RF naturally controls model complexity through random sampling and feature selection. This approach prevents overfitting and enhances generalization performance, providing high prediction accuracy across diverse datasets with outliers [45]. Its ability to effectively handle both categorical and continuous variables makes RF a suitable model for our classification task [46].

2.4. Robust outlier removal algorithms

Previous research indicates that the combined effects of temperature and relative humidity, which are closely linked to bioaerosol viability, play a critical role in determining indoor bioaerosol concentrations. As the joint effect of temperature (T) and relative humidity (RH) becomes more favorable for bioaerosol viability, indoor bioaerosol concentrations are expected to increase. This is demonstrated by the gray circles within the red shaded region in the left panel of Fig. 1a, which represents data points showing a valid trend influenced by T and RH. The red

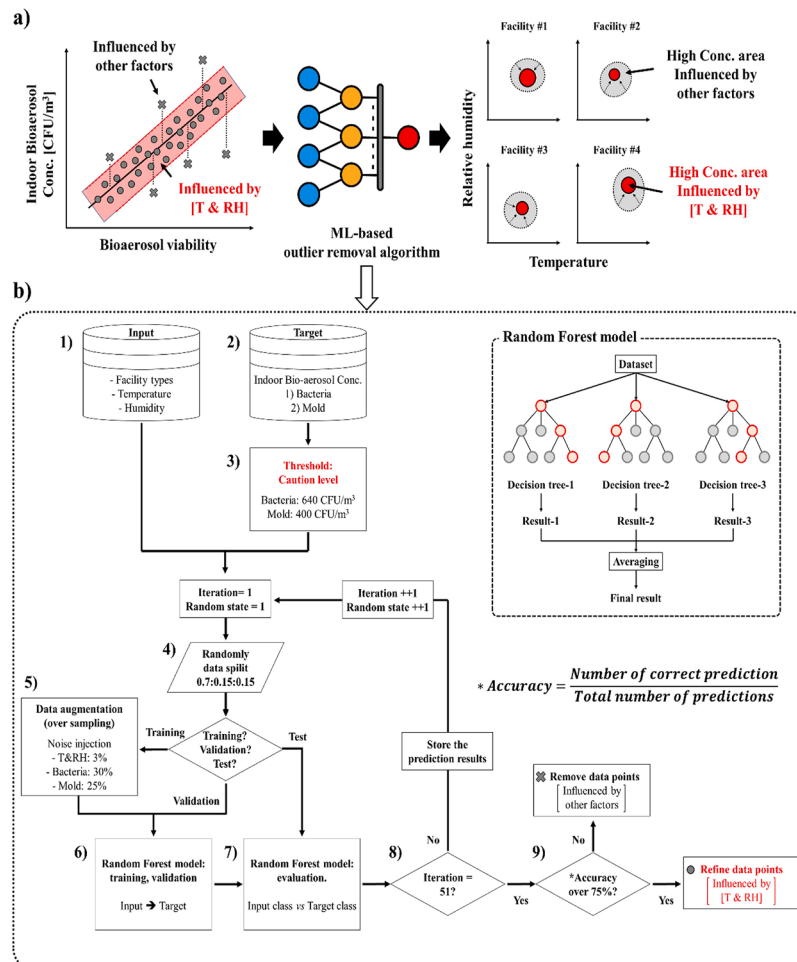


Fig. 1. Machine learning-based outlier removal: a) framework. b) algorithm structure.

shaded region indicates the range of data where bioaerosol viability on the x-axis correlates effectively with indoor bioaerosol concentrations. However, Bioaerosol data collected from real-world environments can be influenced by external factors such as ventilation, occupancy, and building materials. These external factors may result in outliers that deviate from the correlation between viability and indoor bioaerosol concentrations. Such outliers are represented as gray X marks located outside the red-shaded region in the left panel of Fig. 1a. Therefore, to gain a clearer understanding of the relationship between the joint effect of temperature and humidity (i.e. viability) and the concentrations of airborne bacteria and mold indoors, it is essential to remove these outliers.

To eliminate outliers from real-world data, we applied a machine learning (ML)-based outlier removal algorithm, as shown in the middle panel of Fig. 1a. This algorithm analyzes the complex relationships between input variables (facility type, temperature, relative humidity) and the target variable (indoor bioaerosol concentration) to identify and remove outliers, thereby refining the dataset. The right panel of Fig. 1a illustrates the expected outcome of the outlier removal process. Specifically, for each facility group, the gray shaded areas represent the original "high concentration (Caution-Danger) data point regions," which are influenced by both temperature, relative humidity, and other external factors. Once the outliers are removed, the remaining high-concentration data points are concentrated into the red regions, which better reflect the combined effects of temperature (T) and relative humidity (RH). The purpose of the outlier removal process is not to improve accuracy of classification, but rather to extract distributions that reflect only the joint influences of temperature and relative humidity. This allows for a more accurate assessment of how the combined effects of temperature and relative humidity impact indoor bioaerosol concentrations in real-world environments, while minimizing the influence of external factors.

As depicted in Fig. 1b), the detailed ML-based outlier removal algorithm was constructed as follows:

- 1) Set input variables: facility type (label-encoded), indoor temperature, relative humidity
- 2) Set target variable: indoor bioaerosol concentration (bacteria and mold in CFU/m³)
- 3) Define the Caution level thresholds: 640 CFU/m³ for bacteria and 400 CFU/m³ for mold
- 4) Random data split: training/validation/test datasets with a 0.7 ($N = 2834$) / 0.15 ($N = 607$) / 0.15 ($N = 607$) split with a random seed (random state)
- 5) Data augmentation to address the data imbalance within the training dataset
- 6) Training and validation of a random forest model for predicting airborne bacteria and mold concentrations using training and validation datasets
- 7) The prediction accuracy of the Random Forest model trained in step 6 is evaluated using the test dataset. The evaluation compares the class of the measured concentration with the class of the predicted concentration, based on the thresholds defined in step 3, to identify correct and incorrect predictions
- 8) Repeat Steps 4–7 50 times, modifying the random seed (random state) for data splitting in each iteration. This process ensures that prediction results for all data indices in the test dataset are obtained and stored.
- 9) In Step 8, each data index in the test dataset undergoes an average of 15 prediction evaluations. During this process, the number of correct evaluations for each data index is compared to the total number of evaluations. If the accuracy level is below 75 %, the corresponding data index is deemed an outlier and removed, resulting in the creation of a refined dataset

The ML-based outlier removal algorithm identifies and removes

outliers that are determined to be significantly influenced by external factors such as ventilation and occupancy, in addition to temperature and relative humidity, thereby refining the dataset. This allows for a more direct evaluation of the joint effects of temperature and humidity on indoor bioaerosol concentrations. Based on this refined dataset, the study aimed to identify the combined conditions of temperature and humidity conditions that lead to 'Caution level' concentrations of airborne bacteria and mold across different facility types.

3. Results

3.1. Descriptive results

3.1.1. Cautionary level classification

Bacteria and mold concentration data were collected from 4048 indoor locations and categorized into three risk levels according to bioaerosol concentration standards set by the Korean Ministry of the Environment, with safety limits of 800 CFU/m³ for bacteria and 500 CFU/m³ for mold. Concentrations exceeding the standard were classified as 'Danger,' while those below the limit but above 80 % of the standard were classified as 'Caution,' indicating a level that may require attention before reaching the threshold. As shown in Table 1, the concentrations of airborne microorganisms were classified into three categories: Safety, Caution, and Danger, for both bacteria and mold. While only 3.2 % of locations detected bacteria at Caution or Danger levels (≥ 640 CFU/m³), 14.7 % of locations detected mold at caution or danger levels (≥ 400 CFU/m³).

3.1.2. By facility type

Data were collected across four major facility categories: sensitive facilities, transportation-related facilities, temperature-controlled facilities, and other facilities, with approximately 1000 data points gathered for each. Table 2 provides descriptive statistics on airborne bacterial and mold concentrations by facility category. The percentages in parentheses within the table represent the proportion of 'Danger' and 'Caution' level data relative to the total data for each facility type (% Danger/Caution). Higher average airborne bacterial concentrations were observed in daycare centers and underground shopping areas, where ventilation tends to be limited or occupancy is relatively high. Similarly, subway stations and indoor parking lots showed significantly higher mold concentrations, reflecting the ventilation and maintenance challenges in these environments. In daycare centers and underground shopping centers, 14.7 % and 6.8 % of bacteria samples, respectively, were classified as 'Danger' or 'Caution'. For mold samples, 36.1 % in indoor parking lots and 15.7 % in large retail stores fell into the 'Danger' or 'Caution' category.

3.1.3. Seasonality

Fig. 2 illustrates the seasonal variation of bacteria and mold samples classified into the 'Danger' and 'Caution' categories. Each season was defined to span three months for consistency: spring (March–May), summer (June–August), fall (September–November), and winter (December–February). The y-axis in Fig. 2a and 2b represent the number of data points, specifically those corresponding to the 'Caution' and 'Danger' levels, as summarized in Table 1. Notably, the trends for both 'Caution' and 'Danger' levels across different seasons are remarkably similar, reinforcing the validity of our decision to include the 'Caution' category in the analysis. It is also found that non-safe bacteria samples,

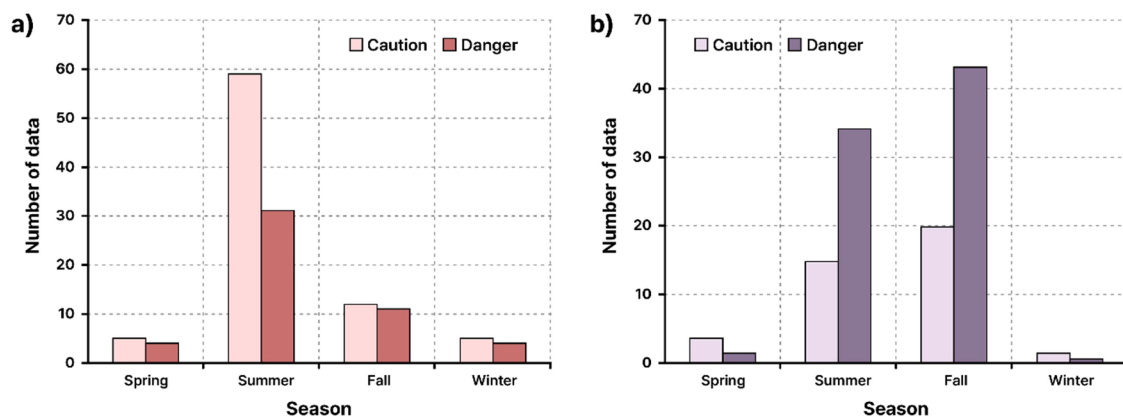
Table 1
Bacteria and mold concentration risk category used in this study.

	Bacteria [CFU/m ³]	mold [CFU/m ³]
Safety	< 640 ($n = 3917$, 96.8 %)	< 400 ($n = 3453$, 85.3 %)
Caution	≥ 640 and < 800 ($n = 81$, 2.0 %)	≥ 400 and < 500 ($n = 198$, 4.9 %)
Danger	≥ 800 ($n = 50$, 1.2 %)	≥ 500 ($n = 397$, 9.8 %)

Table 2

Bacteria and mold concentrations by facility type.

Facility category	Facility type	Bacteria [CFU/m ³] mean $\pm$ SD (% Danger/Caution)	Mold [CFU/m ³] mean $\pm$ SD (% Danger/Caution)	Data counts
Sensitive facilities	Daycare center	377.6 $\pm$ 460.4 (14.7 %)	181.9 $\pm$ 160.9 (10.1 %)	258
	Medical institution	223.6 $\pm$ 179.4 (4.0 %)	133.9 $\pm$ 128.5 (4.6 %)	672
Transportation-related facilities	Bus Terminal	178.0 $\pm$ 158.4 (2.0 %)	151.2 $\pm$ 121.7 (4.0 %)	249
	Underground shopping center	253.3 $\pm$ 219.1 (6.8 %)	162.1 $\pm$ 137.2 (8.6 %)	266
	Subway station	217.5 $\pm$ 188.1 (4.1 %)	221.8 $\pm$ 186.3 (14.6 %)	534
Temperature-controlled facilities	Library	125.9 $\pm$ 158.5 (2.0 %)	181.1 $\pm$ 223.0 (12.5 %)	393
	Museum	97.4 $\pm$ 112.4 (0 %)	102.7 $\pm$ 147.3 (5.2 %)	212
	Movie theater	82.4 $\pm$ 87.1 (0.6 %)	86.1 $\pm$ 96.9 (2.1 %)	326
Other facilities	Large retail store	151.8 $\pm$ 154.5 (2.8 %)	179.9 $\pm$ 227.6 (15.7 %)	249
	Indoor parking lot	107.8 $\pm$ 103.7 (0.4 %)	414.2 $\pm$ 425.0 (36.1 %)	889

**Fig. 2.** Seasonal variation of 'Caution' and 'Danger' samples for; a) bacteria, b) mold.

encompassing both 'Danger' and 'Caution' categories, are predominantly observed during the summer season. In contrast, most 'Danger' and 'Caution' mold samples are found in the summer and autumn seasons. This finding suggests that indoor bioaerosol concentration trends, largely influenced by outdoor air sources, vary with seasonal changes [39]. Bacteria and mold concentrations tend to be higher in summer and autumn, seasons marked by higher temperatures and humidity. Bioaerosol concentrations peak during the hottest, most humid summer months, while the lowest levels occur during the coldest, driest winter months. These seasonal variations in bioaerosol concentrations are therefore key indicators for predicting exceedances of indoor bioaerosol concentration thresholds.

3.1.4. Temperature and relative humidity

To evaluate the effects of temperature and humidity on bacterial and mold concentrations, Pearson correlation coefficients were calculated using all collected data before outlier removal and are presented by facility type in Table 3. This analysis was performed solely for the statistical evaluation of the collected dataset. The analysis revealed a significant impact of these environmental factors on bacteria and mold levels, as variations in temperature and humidity are known to influence bioaerosol viability. Notably, significant positive correlations ($p < 0.01$) were observed in all facility types, with the strongest correlations found in underground shopping centers and subway stations for both temperature and humidity. Relative humidity generally showed a higher correlation with bacterial and mold concentrations compared to temperature, particularly in medical institutions (bacteria: 0.6359) and daycare centers (mold: 0.5490).

3.2. Results of outlier removal using random forest algorithms

For bacteria, the data classified as 'Caution-Danger' level accounted

Table 3

Pearson's correlation coefficients between environmental factors (temperature and relative humidity) and indoor bioaerosol concentrations (bacteria and mold) by facility type.

Facility types	Bacteria		Mold	
	Temperature	Humidity	Temperature	Humidity
Total facility	0.3262*	0.3424*	0.3008*	0.4119*
Daycare center	0.3818*	0.4342*	0.3633*	0.5490*
Medical Institution	0.2961*	0.6359*	0.1914*	0.4721*
Bus terminal	0.5552*	0.4981*	0.4057*	0.3815*
Underground shopping center	0.5580*	0.4250*	0.4560*	0.3815*
Subway station	0.5561*	0.4040*	0.4737*	0.4912*
Library	0.3622*	0.4035*	0.2470*	0.4509*
Museum	0.4085*	0.3572*	0.4210*	0.4721*
Movie theater	0.2098*	0.2252*	0.2379*	0.3596*
Large retail store	0.3908*	0.4527*	0.2178*	0.4053*
Indoor parking lot	0.3512*	0.3215*	0.4341*	0.4501*

* $p < 0.01$.

for only 3.2 % of the total dataset, while for mold, the 'Caution-Danger' level data represented 14.7 % of the total. To address the issue of reduced predictive accuracy for the 'Caution-Danger' level due to this data imbalance, we applied the data augmentation method using over-sampling. This over-sampling technique helped improve the accuracy of the RF-based prediction model by creating a range of environmental conditions for high concentrations of bacteria and mold.

Table 4 shows the results of the RF model performance for classifying 'Safety' and 'Caution-Danger' levels before and after applying the outlier removal algorithm. For airborne bacteria, out of a total of 4048 measured data points, including 3917 'Safety' levels and 131 'Caution-Danger' levels, the RF model using facility type, temperature, and relative humidity as independent variables accurately classified 3876

Table 4

RF model performance for classifying 'safety' and 'caution-danger' levels before and after outlier removal.

	Bacteria			Mold		
	Safety	Caution-Danger	R ²	Safety	Caution-Danger	R ²
Total data	3917	131	0.41	3463	595	0.43
Classified data	3876	37 (28.2 %)	0.52	3173	377 (63.4 %)	0.64
Balanced accuracy	99.0 %			91.9 %		
	63.6 %			77.6 %		

data points as 'Safety' (99.0 %) and 37 data points as 'Caution-Danger' (28.2 %). This resulted in a balanced accuracy of 63.6 % (Table 4). For airborne mold, a total of 4048 data points were measured, including 3453 classified as 'Safe' and 595 as 'Caution-Danger.' The RF model successfully classified 3173 data points as 'Safe' (91.9 %) and 377 data points as 'Caution-Danger' (63.4 %). This resulted in a balanced accuracy of 77.6 % (Table 4). Due to the limited number of data points in the 'Danger' category, it was combined with the 'Caution' level for analysis. ~~Additional results for the 'Danger' category are provided in the supplementary material.~~

We then used an outlier removal algorithm to calculate the prediction accuracy for each index, refining only those datasets with an accuracy of 75 % or higher. In Figs. 3(a) and 3(b), the total data points are represented as black data points, while the refined data points, with outliers removed, are shown as red data points. The black horizontal and vertical lines in each figure indicate the classification criteria for the two classes: 640 CFU/m³ for bacteria in Fig. 3(a) and 400 CFU/m³ for mold in Fig. 3(b). The distribution of red data points relative to these lines highlights the effectiveness of the outlier removal algorithm, ensuring that only the prediction indices with accuracy levels of 75 % or higher are retained. The clear separation of refined data across the class boundaries demonstrates the robustness and validity of the outlier removal process. Once the data points identified as outliers were removed based on the ML-based algorithm, the R² value for the "refined" bacterial concentration data improved to 0.5199, while the R² value for the "refined" mold concentration data increased to 0.6378.

3.3. Joint effect of temperature and relative humidity across all facilities

Fig. 4 visually depicts the distribution of bacterial and mold concentrations in indoor air as the entire dataset and the only "refined" dataset. The left graph presents all data points prior to outlier removal, while the right graph displays the results after applying the outlier removal algorithm proposed in this study, excluding outlier data points. This figure visualizes the distribution areas of bacteria and mold concentrations classified in Safety level (640 CFU/m³ or below; blue area) and Caution-Danger level (above 640 CFU/m³; red area) across temperature-humidity zones for all facilities, separately for the entire

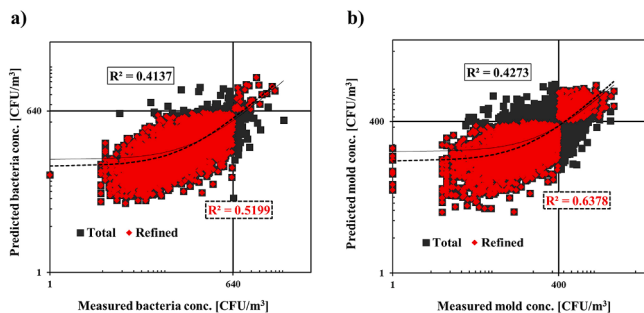


Fig. 3. Scatterplots of measured vs. predicted concentration values: a) bacteria, b) mold.

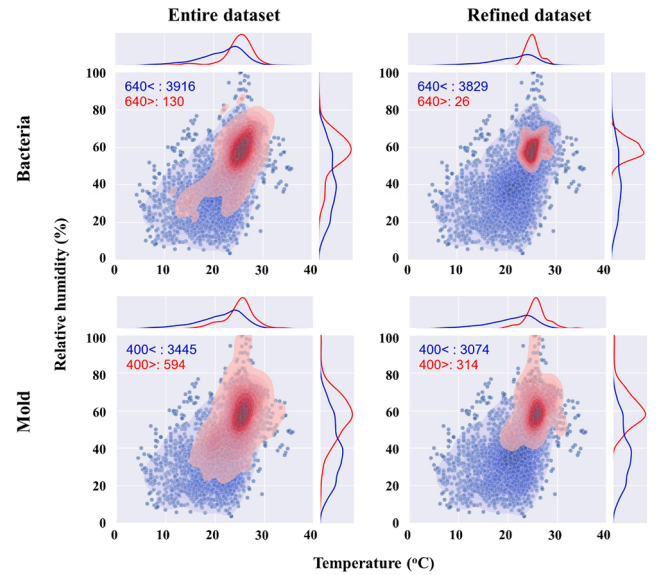


Fig. 4. Joint effect of temperature and relative humidity on bacteria and mold: before vs. after refinement process (all facilities).

dataset and the only "refined" dataset. The numbers within the graph indicate the count of data points, along with the mean and standard deviation of temperature and relative humidity for each category. Additionally, this graph shows the probability density distributions for 'Safety' data points (blue line) and 'Caution-Danger' data points (red line) relative to temperature and relative humidity. Each distribution highlights the ranges of temperature or relative humidity where the data points are most concentrated, particularly 'Caution-Danger' data points. Table 5 provides the numerical details of the frequency distributions shown in Fig. 4, presenting the mean and standard deviation of temperature and relative humidity for both Safety and Caution-Danger data points before and after refinement. These values quantitatively indicate the specific ranges where the highest frequencies occur and illustrate the extent of the distribution for 'Safety' and 'Caution-Danger' data points.

In the bacterial and mold datasets before refinement (left panels of Fig. 4), Safety data points are widely dispersed across a broad range of temperature and relative humidity conditions, while the Caution-Danger data points exhibit more localized clustering patterns, concentrated within specific ranges of temperature and humidity. The quantitative details from Table 5 confirm this observation. For bacteria, Safety data points were distributed at a temperature of 21.7 ± 4.8 °C and a relative humidity of 41.7 ± 16.6 %, whereas Caution-Danger data points were more concentrated at 25.0 ± 3.1 °C and 56.5 ± 11.1 %. Similarly, for mold, Safety data points were observed at 21.2 ± 4.8 °C and 39.7 ± 15.9 %, while the Caution-Danger data clustered at 24.9 ± 3.2 °C and 56.7 ± 13.0 %.

After refinement (right panels of Fig. 4), the Safety data distributions for both bacteria and mold remain largely unchanged, maintaining their broad spread, with only minimal shifts observed. For bacteria, the refined Safety data is distributed at 21.6 ± 4.8 °C and 41.3 ± 16.6 %, while for mold, it stabilizes at 20.8 ± 4.8 °C and 37.5 ± 14.9 %. In contrast, the Caution-Danger data areas show substantial changes after refinement. For bacteria, the Caution-Danger data becomes notably more condensed, clustering tightly at 25.4 ± 1.3 °C and 59.1 ± 4.6 %, with a 58.1 % reduction in the temperature standard deviation and a 58.6 % reduction in the relative humidity standard deviation, leading to an approximate 84 % decrease in the core Caution-Danger area size. For mold, while a similar condensation is observed, the reduction is less pronounced: the Caution-Danger data clusters at 25.7 ± 2.3 °C and 61.0 ± 10.0 %, with a 28.1 % and 23.1 % decrease in the standard deviations of temperature and relative humidity, respectively, resulting in a 45 %

Table 5
Distribution of temperature and relative humidity for bacteria and mold: before and after refinement (all facilities).

		Entire dataset		Refined dataset	
		Safety	Caution-danger	Safety	Caution-danger
Bacteria	Temperature	21.7 ± 4.8	25.0 ± 3.1	21.6 ± 4.8	25.4 ± 1.3
	Relative Humidity	41.7 ± 16.6	56.5 ± 11.1	41.3 ± 16.6	59.1 ± 4.6
Mold	Temperature	21.2 ± 4.8	24.9 ± 3.2	20.8 ± 4.8	25.7 ± 2.3
	Relative Humidity	39.7 ± 15.9	56.7 ± 13	37.5 ± 14.9	61 ± 10.0

reduction in the core area size. These results indicate that bacterial concentrations are more tightly influenced by specific temperature and relative humidity ranges after refinement, whereas mold concentrations retain relatively broader variability even following refinement.

Furthermore, in the 'Caution-Danger' data points, the refinement process notably raised the average temperature and relative humidity by 0.4 °C and 2.6 % for bacteria, and by 0.8 °C and 4.3 % for mold,

respectively. This result suggests that the refinement process could allow us to identify the optimal joint range of temperature and relative humidity for forming actual 'Caution-Danger' levels of indoor bioaerosol (bacteria and mold) concentrations.

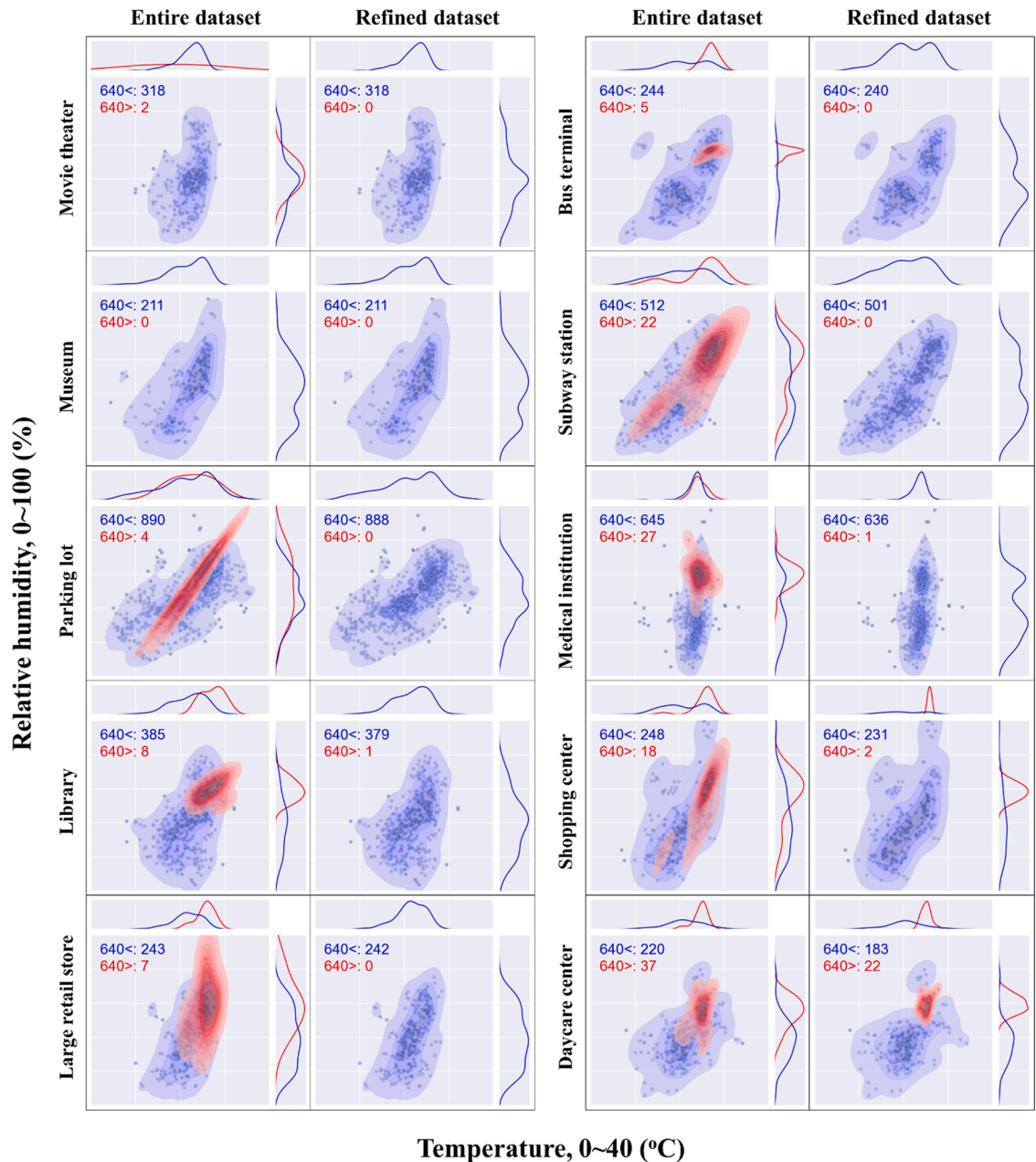


Fig. 5. Joint effect of temperature and relative humidity on bacteria by facility: before vs. after refinement process.

3.4. Joint effect of temperature and relative humidity by facility

Figs. 5 and 6, constructed similarly to Fig. 4, illustrate the distribution of bacteria and mold data by facility type, focusing on changes in the 'Caution-Danger' areas before and after applying the outlier removal algorithm. These figures depict bacterial and mold concentration data categorized into 'Safety' (blue points/areas) and 'Caution-Danger' (red points/areas) levels based on temperature and relative humidity. Section 3.3 confirmed that the 'Safety' area showed minimal changes following refinement, while the 'Caution-Danger' area exhibited significant changes. Therefore, Section 3.4 concentrates on analyzing the 'Caution-Danger' area to better understand the effects of refinement. Tables 6 and 7 complement Figs. 5 and 6 by providing quantitative data on the mean, standard deviation, and refinement rate (RR) for temperature and relative humidity in the 'Caution-Danger' area. The RR represents the proportion of data remaining after outlier removal. High RR values indicate that temperature and relative humidity are the primary factors influencing bacterial or mold concentrations, whereas low RR values suggest that external factors such as occupancy density, ventilation patterns, or outdoor air exchange have a greater impact. In cases where the RR is 0.0 %, no refined dataset remains, implying that the 'Caution-Danger' data in those facilities is predominantly influenced by external factors rather than temperature and relative humidity. In such cases, the algorithm removes all 'Caution-Danger' data, leaving no

refined dataset for further analysis. This analysis of refinement rates (RR) and distribution changes aims to identify facilities capable of effectively managing bacterial and mold concentrations based on the combined distribution of temperature and relative humidity. Furthermore, it seeks to determine high-risk temperature and humidity combinations that lead to elevated bacterial and mold levels, providing valuable insights for targeted management strategies.

For bacteria, Fig. 5 shows that facilities can be classified into two groups: those that retain high-concentration areas after refinement, such as Daycare Centers and Underground Shopping Centers, and those where high-concentration areas are completely removed after refinement. As indicated in Table 6, facilities exceeding the average refinement rate of 9.6 % include Daycare Centers (59.5 %), Underground Shopping Centers (11.1 %), and Libraries (12.5 %). Based on these visual and numerical results, we determined that temperature and relative humidity significantly contribute to airborne bacterial concentrations in Daycare Centers and Underground Shopping Centers. The two facilities displayed highly similar ranges of temperature and relative humidity associated with 'Caution-Danger' after refinement, supporting the relevance of the refinement algorithm.

For mold, Fig. 6 classifies facilities into three groups: those with relatively circular 'Caution-Danger' areas, like Parking Lots and Subway Stations; those with elliptical 'Caution-Danger' areas, such as Museums, Medical Institutions, Underground Shopping Centers, Large Retail

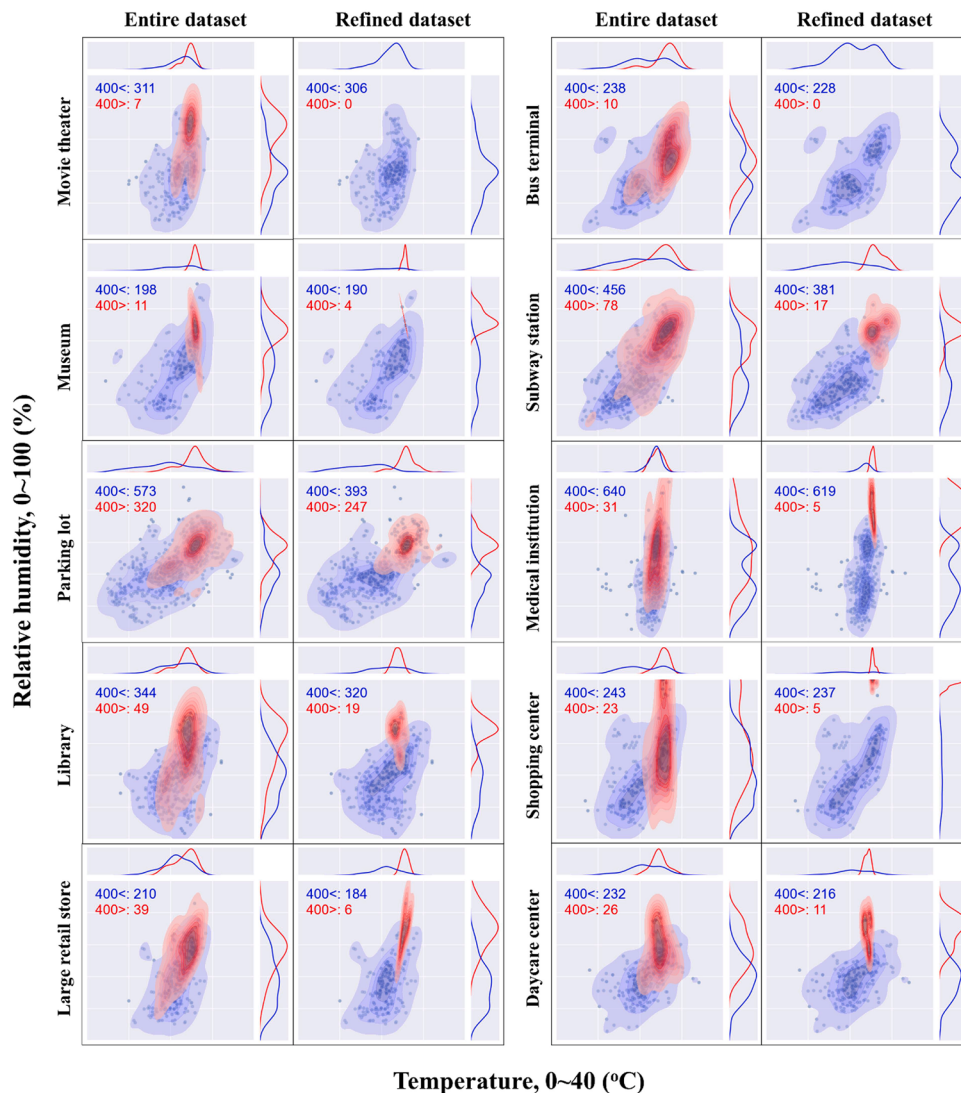


Fig. 6. Joint effect of temperature and relative humidity on mold by facility: before vs. after refinement process.

Table 6

Distribution of temperature and relative humidity for bacteria in 'Caution-Danger' level before and after refinement by facility (RR: refinement rate, RH: relative humidity).

Facility types	RR	Entire Dataset		Refined Dataset	
		Temperature	RH	Temperature	RH
Movie theater	0.0 %	18.6 ± 9.6	42.4 ± 8.8	–	–
Museum	N.A.	No 'Caution-Danger' case		No 'Caution-Danger' case	
Indoor parking lot	0.0 %	22.6 ± 4.7	53.2 ± 17.1	–	–
Library	12.5 %	27.3 ± 2.5	59.1 ± 5.9	28.3 ± 0.1	58.7 ± 0.1
Large retail store	0.0 %	25.7 ± 2	59.5 ± 15.4	–	–
Bus terminal	0.0 %	27 ± 1.6	55.6 ± 2.5	–	–
Subway station	0.0 %	25 ± 5.1	56.1 ± 15.3	–	–
Medical institution	0.4 %	24.6 ± 1.6	59.7 ± 7.2	28.5 ± 0.1	56.8 ± 0.1
Underground shopping center	11.1 %	25.6 ± 2.6	55.5 ± 14.6	25.9 ± 0.4	58.4 ± 4.7
Daycare center	59.5 %	24.7 ± 1.6	54.9 ± 8.4	25.1 ± 1.1	59.2 ± 4.9

Table 7

Distribution of temperature and relative humidity for mold in 'Caution-Danger' level before and after refinement by facility (RR: refinement rate, RH: relative humidity).

Facility types	RR	Entire Dataset		Refined Dataset	
		Temperature	RH	Temperature	RH
Movie theater	0.0 %	24.4 ± 1.3	61.3 ± 13.8	–	–
Museum	36.4 %	25.7 ± 0.7	65.1 ± 10.4	25.5 ± 0.4	73.1 ± 5.5
Indoor parking lot	77.2 %	25.3 ± 3.4	55.8 ± 9.8	25.9 ± 2.4	58.3 ± 7.4
Library	38.8 %	23.5 ± 2.2	57.1 ± 17.1	23.8 ± 1.0	67.4 ± 7.8
Large retail store	15.4 %	23.3 ± 2.6	55 ± 13.5	25.6 ± 0.8	69.7 ± 11.3
Bus terminal	0.0 %	26.3 ± 2.7	49 ± 11.8	–	–
Subway station	21.8 %	24.6 ± 3.8	58.4 ± 14.2	26.7 ± 1.9	66.4 ± 7.7
Medical institution	16.1 %	24.2 ± 1.3	56.7 ± 18.1	25.4 ± 0.4	81.2 ± 9.5
Underground shopping center	21.7 %	25.5 ± 1.5	61.5 ± 24.4	25.6 ± 0.4	98.4 ± 2.5
Daycare center	42.3 %	24.7 ± 1.9	58.5 ± 12.2	24.2 ± 0.7	68.7 ± 9.0

Stores, Libraries, and Daycare Centers; and facilities where 'Caution-Danger' areas were completely removed, like Movie Theaters and Bus Terminals. According to Table 7, facilities exceeding the average refinement rate of 27 % include Parking Lots (77.2 %), Daycare Centers (42.3 %), Libraries (38.8 %), and Museums (36.4 %). Based on these visual and numerical findings, we evaluated Parking Lots as a facility

type where temperature and relative humidity significantly contribute to airborne mold concentrations, while Daycare Centers, Libraries, and Museums were identified as facilities where temperature plays a meaningful role in influencing airborne mold concentrations. Fig. 7 illustrates refinement rate for bacteria and mold by facility.

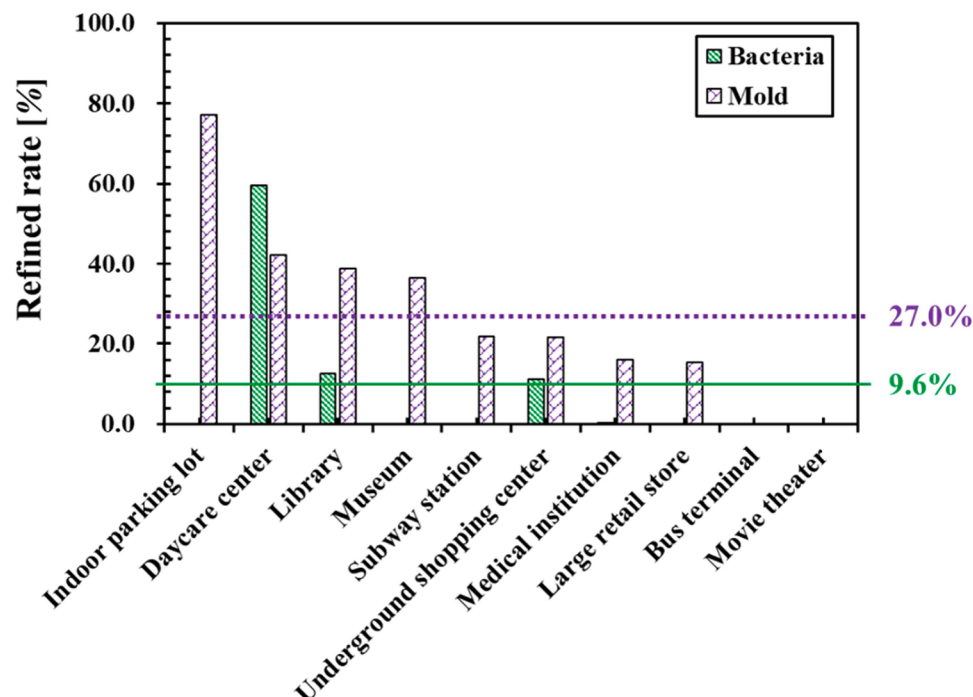


Fig. 7. Refinement rate for bacteria and mold by facility.

4. Discussion

This study collected 212 to 889 data points per facility over two years, covering all four seasons, to establish a diverse set of environmental conditions that include a broader range of temperature and relative humidity as well as various density conditions within indoor spaces. Through this, it was possible to identify the joint effects of temperature and relative humidity on indoor bioaerosol concentrations—effects not previously observed in laboratory-based or field studies. Specifically, the machine learning-based refinement algorithm proposed in this study filtered out noise data points affected by extraneous factors, allowing for a clearer and more direct observation of the joint effects of temperature and relative humidity. In fact, it has been reported that temperature and relative humidity generally show a positive correlation with indoor airborne bacterial concentrations [47–50], which aligns with this study's finding that, after applying the refinement algorithm, the joint range of temperature and humidity associated with high bacterial concentrations ('Caution-Danger' data points) became more concentrated. Furthermore, mold has a greater ability to adapt to diverse environmental conditions compared to bacteria, enabling it to survive and proliferate across a broader range of temperatures and humidity levels, thereby sustaining high concentrations [51,52]. The finding in this study that mold exhibits a relatively wider 'Caution-Danger' range of temperature and humidity compared to bacteria is consistent with previous research. These findings indicate that when government and institutions develop indoor air quality management strategies, bacterial levels should be carefully controlled and monitored within specific temperature and humidity ranges, whereas mold management requires broader oversight across a wider spectrum of conditions, reflecting its adaptability to diverse environmental factors. An additional contribution of this study is its provision of evidence for classifying different facility types based on the joint effects of temperature and relative humidity on airborne bacterial viability, facilitating targeted strategies for each group. However, indoor air quality management involves more than just temperature and humidity control; factors such as seasonality, building materials, and occupant comfort must also be considered. In regions with distinct seasonal variations, such as Korea, where cold winters and snow impose additional constraints, temperature and humidity control should be approached with greater caution to ensure effective and comprehensive IAQ management.

In Sections 3.3 and 3.4, it was observed that the outliers removed during the refinement process were predominantly concentrated in the high bioaerosol concentration range. This suggests that temperature and relative humidity have a more pronounced impact on the reduction (decay) term of bioaerosol survival rather than on the generation term. Specifically, when temperature and relative humidity deviate from ranges that support high bioaerosol viability, the decay factors become more dominant. Conversely, high indoor bioaerosol concentrations are likely influenced more by generation factors, such as occupancy density and outdoor air exchange, rather than by the decay effects of temperature and relative humidity. Thus, these outliers primarily reflect generation-driven influences rather than the combined effects of temperature and relative humidity. Nevertheless, the refined dataset is considered to effectively capture real-world conditions. For instance, the removal of 'Caution-Danger' data in certain facility groups does not imply that high bioaerosol concentrations cannot occur in these settings. Rather, it indicates that high concentrations are less likely to result from the combined effects of temperature and relative humidity in such facilities. However, factors such as increased occupancy or outdoor air intrusion could still contribute to elevated bioaerosol levels. In conclusion, the refined dataset enables a more accurate analysis of the isolated effects of temperature and relative humidity while acknowledging the potential impact of external factors, thereby maintaining its applicability to real-world scenarios.

For bacteria, in facility types such as cinemas, museums, indoor

parking garages, libraries, large retail stores, bus terminals, subway stations, and healthcare facilities, the removal of 'Caution-Danger' data points during the refinement process indicates that external factors may have a greater influence on indoor bioaerosol concentrations than temperature and relative humidity. For instance, facilities such as cinemas, museums, and libraries strive to maintain stable indoor environments but are influenced by external factors like occupancy density, ventilation levels, and outdoor air exchange. As a result, changes in human activity and airflow may have a greater impact on indoor bioaerosol concentrations than temperature or relative humidity. Similarly, high-occupancy facilities with limited ventilation, such as underground shopping malls, subway stations, and bus terminals, may be more affected by population density and air circulation patterns than by temperature and humidity [34,53]. In healthcare facilities, temperature and relative humidity are generally kept stable; however, due to the presence of various pathogens and the use of air purification and ventilation systems, bioaerosol concentrations may rely more on external management factors than on environmental stability [54,55]. Facility groups such as daycare centers and underground shopping complexes, where "Caution-Danger" data appears as circular clusters after applying the refined model, show that temperature and relative humidity operate within narrow, complementary ranges that significantly impact indoor bioaerosol concentrations. In daycare centers, where stringent cleanliness is essential, both temperature and humidity are carefully controlled within these tight parameters. To maintain these specific conditions, air exchange with the outside is minimized, reducing the influence of external factors and making temperature and relative humidity key determinants of the indoor environment [35–37].

Regarding mold, facilities with a significant refinement rate were categorized into three groups based on differential impacts of temperature and relative humidity. First, in facility groups such as movie theaters and bus terminals, where "Caution-Danger" data was entirely filtered out during the refinement process, external factors may more strongly influence indoor airborne mold concentrations than temperature or humidity. Movie theaters aim to maintain a stable indoor environment but are still impacted by external factors like occupancy density, ventilation, and air exchange with the outdoors. In such settings, human activity and changes in air flow likely have a greater effect on indoor bioaerosol concentrations than temperature or humidity. Likewise, bus terminals, with their high occupancy density and restricted ventilation, are influenced more by population density and air circulation patterns than by temperature and humidity. Second, facilities such as museums, medical institutions, underground shopping centers, libraries, and daycare centers—where high mold concentrations form in an elliptical pattern—could tend to have a broader range of relative humidity but a narrower temperature range. This pattern suggests that temperature is a more influential factor than humidity in promoting high indoor airborne mold concentrations. In museums and libraries, for instance, temperature is tightly controlled to preserve artifacts and books, while humidity regulation is more challenging, resulting in greater variation [56]. A relative humidity below 60 % is recommended to prevent high mold concentrations in museums [57], which aligns with the humidity levels associated with "Caution-Danger" points for museums in this study. In medical facilities, temperature is kept relatively stable to ensure patient comfort and operational efficiency, but relative humidity can vary depending on the effectiveness of ventilation and air conditioning systems. Previous studies indicate that mold control in healthcare settings remains challenging due to the difficulties in humidity regulation [58,59]. The presence of high mold concentrations even under wide-ranging humidity conditions in these facilities suggests that mold can thrive across diverse humidity levels, implying that relative humidity may not be a critical control factor for mold concentration [51,52]. Lastly, in facility groups such as indoor parking lots and subway stations, where "Caution-Danger" mold data appears circular after refinement, temperature and relative humidity jointly impact indoor airborne mold concentrations within narrow

ranges. Both types of facilities are underground, limiting exposure to external weather fluctuations and reducing air circulation from open building areas. Consequently, the interaction between indoor temperature and relative humidity becomes crucial in influencing airborne mold concentrations, outweighing the effects of external factors [60,61].

This study has four notable limitations that suggest directions for future research. First, while we developed a refinement algorithm to remove data points likely influenced by external factors, allowing for a clearer analysis of how temperature and relative humidity affect indoor bioaerosol concentrations, we were unable to validate these assumptions using actual data on external conditions. To enhance the accuracy and reliability of this analysis, future research should collect data on specific external conditions for each facility type to enable a more comprehensive understanding of these influences. Additionally, the role of building materials and indoor decorations, such as carpets, curtains, and upholstered furniture, should be considered as potential sources of mold. Facilities like theaters, which often feature such materials, may accumulate mold that becomes aerosolized irrespective of temperature and relative humidity conditions. Including these factors in future studies would provide a more holistic understanding of mold sources and their contribution to indoor bioaerosol levels. Moreover, environmental variables, such as building issues (e.g., fires, leaks) or extreme weather events (e.g., heat waves, cold snaps), may alter the impact of temperature and humidity on indoor bioaerosol levels. Collecting long-term data across diverse environmental conditions would therefore be beneficial.

Second, this study employs an outlier removal algorithm based on a Random Forest (RF) model to support nonlinear analysis and improve generalization across various conditions. However, outlier detection can vary depending on the RF model's parameters and structure, and alternative models or optimization methods could outperform RF in certain cases. Future studies should explore multiple models and parameter optimizations to identify a more effective refinement approach. Additionally, it is important to acknowledge that our outlier removal algorithm may not entirely isolate the pure joint effect of temperature and humidity from external influences.

Third, while this study evaluates overall concentrations of bacteria and mold, each microorganism has specific optimal conditions for growth (e.g., temperature and relative humidity), which may not be fully generalized across all microorganisms. Additionally, only viable and culturable bacteria and mold were sampled, meaning that the study may underestimate actual concentrations, particularly for microorganisms that are non-culturable under the given conditions. These limitations highlight the importance of interpreting the results with caution and suggest that future research should consider incorporating molecular techniques or other methods that account for non-culturable microorganisms.

Finally, to enhance the practical application of this study's findings, more field-based experiments are needed to assess how temperature and humidity affect "Caution-Danger" points of bioaerosol concentrations across various scenarios that consider interactions with external variables unique to each facility type. Testing the effectiveness of different strategies under diverse conditions is crucial for developing targeted, effective approaches to improve indoor air quality in a range of facility environments. Additionally, although this study collected a comprehensive dataset, the number of usable data points for certain "Caution-Danger" distributions was relatively small after refinement for some facility types. For example, in facility types such as movie theaters and bus terminals, the refined dataset was limited to 4–8 data points, which may constrain the robustness of conclusions regarding specific temperature and humidity distributions in these facilities. While the conclusions align with the context of each facility and are compelling, it is essential to acknowledge the limited data for certain facility types and interpret the findings with caution. Future research should focus on collecting additional data points for these facility types to enhance statistical reliability and enable more precise characterization of

temperature and humidity distributions.

5. Conclusions

This study comprehensively analyzes the effects of temperature, relative humidity, and external factors on indoor bioaerosol concentrations, clustering various facilities based on these factors. It clearly illustrates the relative importance of temperature, relative humidity, and external influences according to the environmental characteristics of each facility group. The classification of these facilities proposed in this study offers important implications for indoor air quality management. For facility groups like theaters and bus terminals, where external factors exert a greater influence, strategies focused on improving ventilation and managing occupancy levels are essential. For facilities such as museums and libraries, where controlling fluctuations in relative humidity is challenging, it would be more practical and effective to prioritize temperature as the primary control variable to prevent the optimization of mold concentrations. In environments like daycare centers, indoor parking lots, and subway stations, where there is limited air exchange with the outside, it is crucial to manage both indoor temperature and relative humidity in balance. This research underscores the significance of implementing customized indoor environment management strategies tailored to the unique characteristics of each facility and offers insights into how even non-experts can manage indoor bioaerosol concentrations by simultaneously adjusting temperature and relative humidity, which are relatively easier to control.

CRediT authorship contribution statement

Doheon Kim: Writing – original draft, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Dongmin Shin:** Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation. **Dohyeong Kim:** Writing – review & editing, Supervision, Software, Resources, Methodology, Formal analysis, Conceptualization. **Boyeon Kwon:** Writing – original draft, Methodology, Formal analysis, Data curation. **Choongki Min:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Gloria Geervarghese:** Data curation, Formal analysis, Writing – review & editing. **Seunghyun Kim:** Writing – review & editing, Visualization, Methodology, Data curation. **Jungho Hwang:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **SungChul Seo:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Korea Environment Industry and Technology Institute (KEITI) through the Technology Development Project for Biological Hazards Management in Indoor Air, funded by the Korea Ministry of Environment (MOE) (Grant number 2021003380004 and 2021003370005).

Data availability

The authors do not have permission to share data.

References

- [1] E. Meltzer, D. Bukstein, The economic impact of allergic rhinitis and current guidelines for treatment, *Annal. Allergy, Asthma Immunol.* 106 (2011) S12–S16.
- [2] S. Walser, D. Gerstner, B. Brenner, J. Bungler, T. Dikmann, B. Janssen, et al., Evaluation of exposure–response relationships for health effects of microbial bioaerosols – a systematic review, *Int. J. Hyg. Environ. Health* 218 (2015) 577–589.
- [3] S. Alnahas, N. Abouammoh, W. Althagafi, E. Abd-Ellatif, Prevalence, severity, and risk factors of allergic rhinitis among schoolchildren in Saudi Arabia: a national cross-sectional study, 2019, *World Allergy Org. J.* 16 (2023) 100824.
- [4] N. Karakousis, O. Kotsiou, K. Gourgoulis, Bronchial asthma and sarcopenia: an upcoming potential interaction, *J. Pers. Med.* 12 (2022) 1556.
- [5] L. Deng, Q. Deng, The basic roles of indoor plants in human health and comfort, *Environ. Sci. Pollut. Res.* 25 (2018) 36087–36101.
- [6] Y. Shan, J. Guo, W. Fan, H. Li, H. Wu, Y. Song, et al., Modern urbanization has reshaped the bacterial microbiome profiles of house dust in domestic environments, *World Allergy Org. J.* 13 (2020) 100452.
- [7] R.A. Samson, B. Flannigan, M.E. Flannigan, A.P. Verhoeff, O.C.G. Adan, E. S. Hoekstra, Health Implications of Fungi in Indoor Environments, Elsevier Science, Amsterdam, Netherlands, 1995.
- [8] S. Seo, Health effects of exposure to indoor mold and the levels of mold in facilities with susceptible populations in Korea, *J. Environ. Health Sci.* 46 (2020) 359–367.
- [9] J. Ha, S.W. Lee, D.K. Yon, Ten-year trends and prevalence of asthma, allergic rhinitis, and atopic dermatitis among the Korean population, *Clin. Exp. Pediatr.* 63 (2020) 278–283.
- [10] E. Lee, A. Kim, Y.M. Ye, S.E. Choi, H.S. Park, Increasing prevalence and mortality of asthma with age in Korea, 2002–2015: a nationwide, population-based study, *Allergy Asthma Immunol. Res.* 12 (2020) 467–484.
- [11] S. Choi, H. Moon, M. Kim, A study on the improvement of indoor air quality system in public facilities in Seoul - after the revision of the indoor air quality control act in 2015, *J. Korean Soc. Living Environ. Syst.* 30 (2022) 50–59.
- [12] H. Lee, B. Kwon, H. Jeong, K. Jo, W. Yoon, Analysis of the status of suspended fungal concentration in indoor air of national multi-use facilities and their correlation with characteristics and contributing factors according to regional and facility classifications, *J. Odor Indoor Environ.* 21 (2022) 170–180.
- [13] Y. Cho, H. Kim, H. Ko, S. Jeong, B. Kim, J. Jung, Continuous surveillance of bioaerosols on-site using an automated bioaerosol-monitoring system, *ACS. Sens.* 5 (2020) 395–403.
- [14] J. Park, H. Kim, J. Hwang, Continuous and real-time bioaerosol monitoring by combined aerosol-to-hydrosol sampling and ATP bioluminescence assay, *Anal. Chim. Acta* 941 (2016) 101–107.
- [15] A. Paton, J. Paton, A. Lawrence, P. Goldwater, R. Harris, Rapid detection of respiratory syncytial virus in nasopharyngeal aspirates by reverse transcription and polymerase chain reaction amplification, *J. Clin. Microbiol.* 30 (1992) 901–904.
- [16] M.F. Clark, A.N. Adams, Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses, *J. General Virol.* 34 (1977) 475–483.
- [17] R. Chandra, S. Goyal, R. Gupta, Evaluation of deep learning models for multi-step ahead time series prediction, *IEEe Access.* 9 (2021) 83105–83123.
- [18] U. Pak, J. M. U. Ryu, K. Ryou, U. Juhuyok, K. Pak, et al., Deep learning-based PM_{2.5} prediction considering the spatiotemporal correlations: a case study of Beijing, China, *Sci. Total Environ.* (2020) 699.
- [19] R. Wathore, S. Rawlekar, S. Anjum, A. Gupta, H. Bherwani, N. Labhasetwar, et al., Improving performance of deep learning predictive models for COVID-19 by incorporating environmental parameters, *Gondwana Res.* 114 (2023) 69–77.
- [20] W.H. Finch, Distribution of variables by method of outlier detection, *Front. Psychol.* 3 (2012).
- [21] Rahmatullah Imon AHM, Identifying multiple influential observations in linear regression, *J. Appl. Stat.* 32 (2005) 929–946.
- [22] J. Tolvi, Genetic algorithms for outlier detection and variable selection in linear regression models, *Soft comput.* 8 (2004) 527–533.
- [23] A. Micheluz, S. Manente, V. Tigini, V. Prigione, F. Pinzari, G. Ravagnan, et al., The extreme environment of a library: xerophilic fungi inhabiting indoor niches, *Int. Biodeterior. Biodegradat.* 99 (2015) 1–7.
- [24] R. Ehrlich, S.O. Miller, R.L. Walker, Relationship between atmospheric temperature and survival of airborne bacteria, *Appl. Microbiol.* 19 (1970) 245–249.
- [25] T.W. Ng, W.L. Chan, K.M. Lai, Importance of stress-response genes to the survival of airborne *Escherichia coli* under different levels of relative humidity, *AMB Express.* 7 (2017) 1–6.
- [26] C.S. Cox, Airborne bacteria and viruses, *Sci Prog* 73 (1989) 469–499.
- [27] W. Smets, S. Moretti, S. Denys, S. Lebeer, Airborne bacteria in the atmosphere: presence, purpose, and potential, *Atmos. Environ.* 139 (2016) 214–221.
- [28] J.W. Tang, The effect of environmental parameters on the survival of airborne infectious agents, *J. R. Soc. Interface* 6 (2009) S737–S746.
- [29] S.J. Webb, Factors affecting the viability of air-borne bacteria: I. Bacteria aerosolized from distilled water, *Can. J. Microbiol.* 5 (1959) 649–669.
- [30] G. Jiang, J. Ma, C. Wang, Y. Wang, A. Laghari, Kinetics and mechanism analysis on self-decay of airborne bacteria: biological and physical decay under different temperature, *Sci. Total Environ.* 832 (2022).
- [31] C.M. Wathes, K. Howard, A.J.F. Webster, The survival of *Escherichia coli* in an aerosol at air temperatures of 15 and 30°C and a range of humidities, *J. Hyg.* 97 (1986) 489–496.
- [32] A.L. Pasanen, P. Kallioikoski, P. Pasanen, M.J. Jantunen, A. Nevalainen, Laboratory studies on the relationship between fungal growth and atmospheric temperature and humidity, *Environ. Int.* 17 (1991) 225–228.
- [33] H. Wu, J.W.C. Wong, Temperature versus Relative Humidity: which Is More Important for Indoor Mold Prevention? *J. Fungi* 8 (2022).
- [34] S.H. Hwang, J.B. Park, Comparison of culturable airborne bacteria and related environmental factors at underground subway stations between 2006 and 2013, *Atmos. Environ.* 84 (2014) 289–293.
- [35] R. Jabeen, M.I. Kizhisseri, S.N. Mayanaik, M.M. Mohamed, Bioaerosol assessment in indoor and outdoor environments: a case study from India, *Sci. Rep.* 13 (2023) 18066.
- [36] J. Madureira, I. Paciencia, J.C. Rufo, C. Pereira, J.P. Teixeira, Oliveira de, E. Fernandes, Assessment and determinants of airborne bacterial and fungal concentrations in different indoor environments: homes, child day-care centres, primary schools and elderly care centres, *Atmos. Environ.* 109 (2015) 139–146.
- [37] M.S. Zuraimi, K.W. Tham, Indoor air quality and its determinants in tropical child care centers, *Atmos. Environ.* 42 (2008) 2225–2239.
- [38] R.I. Adams, S. Bhangar, W. Pasut, E.A. Arens, J.W. Taylor, S.E. Lindow, et al., Chamber bioaerosol study: outdoor air and human occupants as sources of indoor airborne microbes, *PLoS. One* 10 (2015) e0128022.
- [39] A. Prussin, L. Marr, Sources of airborne microorganisms in the built environment, *Microbiome* 3 (2015).
- [40] B.K. Iwana, S. Uchida, An empirical survey of data augmentation for time series classification with neural networks, *PLoS. One* 16 (2021) e0254841.
- [41] Wen Q., Sun L., Yang F., Song X., Gao J., Wang X., et al. Time series data augmentation for deep learning: a survey. *arXiv:2002.12478*. 2020.
- [42] Wen T., Keyes R. Time series anomaly detection using convolutional neural networks and transfer learning. *arXiv:1905.13628* 2019.
- [43] K. Kirasich, T. Smith, B. Sadler, Random forest vs logistic regression: binary classification for heterogeneous datasets, *SMU data Sci. Rev.* 1 (2018) 9.
- [44] M. Daho, M. Chikh, Combining bootstrapping samples, random subspaces and random forests to build classifiers, *J. Med. Imag. Health Inform.* 5 (2015) 539–544.
- [45] M. Safaei, M. Driss, W. Boulila, E.A. Sunararajan, M. Safaei, Global outliers detection in wireless sensor networks: a novel approach integrating time-series analysis, entropy, and random forest-based classification, *Software: Pract. Exp.* 52 (2022) 277–295.
- [46] T. Zhu, Analysis on the applicability of the random forest, *Journal of Physics: Conference Series* (2020) 1607.
- [47] H. Avdogdu, A. Asan, M.T. Otkun, M. Ture, Monitoring of fungi and bacteria in the indoor air of primary schools in Edirne city, Turkey, *Indoor Built Environ.* 14 (2005) 411–425.
- [48] B. Kumar, G.P. Gupta, S. Singh, U.C. Kulshrestha, Study of abundance and characterization of culturable bioaerosol at Delhi, India, *Int. J. Environ. Eng. Manag.* 4 (2013) 219–226.
- [49] J.S. Pastuszka, U.K.T. Paw, D.O. Lis, A. Wlazlo, K. Ulfig, Bacterial and fungal aerosol in indoor environment in Upper Silesia, Poland, *Atmos. Environ.* 34 (2000) 3833–3842.
- [50] L.T. Wong, W.Y. Chan, K.W. Mui, P.S. Hui, An assessment of airborne fungi exposure risk level in air-conditioned offices, *Indoor Built Environ.* 18 (2009) 553–561.
- [51] J.H. Lee, W.K. Jo, Characteristics of indoor and outdoor bioaerosols at Korean high-rise apartment buildings, *Environ. Res.* 101 (2006) 11–17.
- [52] C. Xu, H. Chen, Z. Liu, G. Sui, D. Li, H. Kan, et al., The decay of airborne bacteria and fungi in a constant temperature and humidity test chamber, *Environ. Int.* 157 (2021) 106816.
- [53] S. Hwang, W. Park, Radon and PM₁₀ concentrations in underground parking lots and subway stations with health risks in South Korea, *Environ. Sci. Pollut. Res.* 25 (2018) 35242–35248.
- [54] R. Dai, S. Liu, Q. Li, H. Wu, L. Wu, C. Ji, A systematic review and meta-analysis of indoor bioaerosols in hospitals: the influence of heating, ventilation, and air conditioning, *PLoS. One* 16 (2021) e0259996.
- [55] M.S. Mousavi, M. Hadei, M. Majlesi, P.K. Hopke, M. Yarahmadi, B. Eman, et al., Investigating the effect of several factors on concentrations of bioaerosols in a well-ventilated hospital environment, *Environ. Monit. Assess.* 191 (2019) 1–11.
- [56] M. Baldan, S. Manente, F.C. Izzo, The role of bio-pollutants in the indoor air quality of old museum buildings: artworks biodeterioration as preview of human diseases, *Environ. Monit. Assess.* 193 (2021) 1–13.
- [57] C. Bastholm, A. Madsen, B. Andersen, J. Frisvad, J. Richter, The mysterious mould outbreak - a comprehensive fungal colonisation in a climate-controlled museum repository challenges the environmental guidelines for heritage collections, *J. Cult. Herit.* 55 (2022) 78–87.
- [58] J.A. Belizario, L.G. Lopes, R.H. Pires, Fungi in the indoor air of critical hospital areas: a review, *Aerobiologia (Bologna)* 37 (2021) 379–394.
- [59] G. Caggiano, C. Napoli, C. Coretti, G. Lovero, G. Scarafale, O. De Giglio, et al., Mold contamination in a controlled hospital environment: a 3-year surveillance in southern Italy, *BMC Infect. Dis.* 14 (2014) 1–5.
- [60] S.H. Hwang, C.S. Yoon, K.N. Ryu, S.Y. Paik, J.H. Cho, Assessment of airborne environmental bacteria and related factors in 25 underground railway stations in Seoul, Korea, *Atmos. Environ.* 44 (2010) 1658–1662.
- [61] Y. Xing, Y. Li, K. Zhang, D. Liu, G. Zhang, L.Y. Zhao, On-site investigation of the concentrations and size distributions of bioaerosols in the underground garages, *Atmospheric Pollut. Res.* 14 (2023) 101809.